DOCUMENT RESUME

ED 266 967 SE 046 479

AUTHOR Borst, Richard Alan
TITLE A Validity and Reliability Study of a Computer Simulation Designed to Test for Control of Variables.
PUB DATE 86
NOTE 12p.; Paper presented at the Annual Meeting of the National Association for Research in Science Teaching (59th, San Francisco, CA, March 28-April 1, 1986). Several pages are marginally legible due to light type.
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Cognitive Processes; *Computer Simulation; Piagetian Theory; *Reliability; *Science Education; Secondary Education; Secondary School Science; *Validity
IDENTIFIERS Piagetian Tasks; Science Education Research

ABSTRACT

The problem of this study was to develop, validate, and establish the reliability of a computer simulation which would: (1) distinguish between concrete and formal thinking patterns related to control of variables; (2) be easily administered to a class of secondary school age subjects; (3) preserve the interactive aspects of the Piagetian task; (4) be scored using a related computer program; and (5) preserve the questioning of the Piagetian tasks. To accomplish these objectives, 75 secondary school students (38 males and 37 females) were tested using these four instruments: the computer simulation; Piaget and Inhelder's pendulum task; Lawson's Classroom Test of Formal Thinking; and the American Chemical Society Test of High School Chemistry. In addition, a panel of experts evaluated the consistency of the computer simulation with the pendulum task. Based on this evaluation and the statistical relationships between the simulation and pendulum task, it may be concluded that the technique has a degree of construct validity with the pendulum task. Concurrent validity can be established on the basis of the non-parametric correlations examined. Results also indicate a high degree of test-retest reliability (r=0.81). Sex bias was also demonstrated (based on the difference in variance of male and female simulation scores). (Author/JN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

A Validity and Reliability Study of A Computer Simulation Designed to Test for Control of Variables

ED266967

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Presented by

Richard Alan Borst
South Gwinnett High School
2288 East Main Street
Snellville, GA 30278

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Richard Alan Borst

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

SE 046 479

At the NATIONAL ASSOCIATION FOR RESEARCH IN SCIENCE TEACHING; San Francisco, California, March 28, 1986

Any application of Piaget's theory of cognitive development necessitates the development of valid and reliable measures of cognitive level. Inhelder and Piaget's(1958) interview tasks have been adapted for classroom use(Dale, 1970;Sommerville, 1974); however, such instruments require a trained interviewer, equipment and are inconvenient to administer in most classroom situations. Another approach has been to develop paper-and-pencil instruments (Raven 1973,977;Tschopp & Kurdek,1981;Walker, Hendrix, & Merten, 1979; Gray, 1973;Karplus & Karplus, 1970;Staver & Grabel, 1979;Longeot, 1965). Although these instruments are the most convenient for the researcher, written questions may not duplicate the interactive nature of the interview. A third approach has been the development of "demonstration" tasks in which the subject is asked written questions following the presentation of a task with some equipment.(Lawson, 1978;Shayer, Adey & Wylam, 1974). These trends indicate that the development of an instrument involves a trade-off between ease in administration and the preservation of the interactive aspects of Piaget's task. The problem then is one of finding a format which can achieve both.

A little used format for Piagetian based instruments is the microcomputer (Bastien & Bovet,1975). Computerized tasks offer many advantages over the traditionally used formats. A task can be programmed so that it not only preserves the interactive nature of the original measure, but also is easy to administer. Even the questioning involved in the administration of the tasks can be preserved by randomnly generating questions. The scoring of the task can also be computerized to facilitate objectivity and reduce time used in scoring.

Problem Statement

The problem of this study was to develop ,validate, and establish the reliability of a computer simulation which would (1) distinguish between concrete and formal thinking patterns for the construct of control of variables, (2) be easily administered to a class of secondary school age subjects;(3) preserve the interactive aspect of the Piagetian task;(4) be scored using a related computer program and (5) preserve the questioning of the Piagetian tasks.

Development of the Instrument

The instrument under consideration in this study was developed from a commonly used simulation of the Haber process for the production of ammonia. Students input information about the temperature, the pressure and the nature of a catalyst. The program, see Figure 1, calculates a percent yield of ammonia and reports this data on the video monitor. The student is also informed if the pressure is too high; A meltdown will occur. or the pressure is too high;An explosion will occur. In this system an increase in pressure and a low temperature will increase the yield, while the catalyst has no effect.

This simulation is roughly analogous to Inhelder and Piaget's pendulum task. In the pendulum task the student is asked to find the factor of three which will increase the oscillation of the pendulum. To do this he/she manipulates an

BEST COPY AVAILABLE

actual pendulum in which the mass of the bob, the length of the string and the point of release can be varied. In both the chemical simulation and the pendulum task the student must perform a series of experiments in order to exclude the inoperant variables and thereby find the operant one. Theoretically the developmental level of the student is manifested in the way in which the subject goes about doing this.

In order to make the simulation more consistent with Piagetian testing practice, additional aspects were added. Inhelder and Piaget (1958) made use of questioning during the interview to clarity subjects' responses and to have the subjects predict the intent of subsequent trials. This aspect was incorporated in the computer simulation by the use of randomly generated questions about the intent of the next trial. These questions appear three times during the running of the program. Inhelder and Piaget also used questions to ascertain the subjects ability to make conclusions about the effect of each variable in the Pendulum Task. To incorporate this aspect the computer simulation was programmed to present three questions about the effect of each variable. These questions appear after the termination of the experimental program.

The data generated in during the experimental trials involving the Haber process and the answers to the questions are used to calculate a total score for the subject. A separate scoring program was devised to electronically evaluate the data. Five aspects of Sommerville's (1974) quantification of the pendulum task were used to score the data;

1) Formal subjects will only change one variable per trial.

2) Formal subjects will change the correct variable according to their prediction.

3) Formal subjects will change ONLY the predicted variable.

4) Formal subjects will make the correct inferences from the data generated.

5) Formal subjects will make some untested inferences.

The scoring program evaluates the data and answers to the randomly generated and terminal questions. In the first case it calculates a score on the basis of the percentage of the trials which exhibit only one variable change per trial. The ability of the subject to change variables as predicted is evaluated in the second subscore. Since three questions were asked the program computes a score of 33%, 66% or 100% on the basis of the data. In the third and related case a score of 33%,66% or 100% was computed on the basis of the number of trials in which only the predicted variable was changed. In the last two cases the answers to the terminal questions are evaluated along with the data generated. In this case it was considered advantageous to have all conditions of a particular variable represented in the data. In addition the answer to the question about the effect of this variable must be correct to obtain a score in this category. The fifth category about the presence of untested inferences was evaluated in a similar manner. In this case the subject would score this category correctly if he/she had made a correct answer to a terminal question and only have

BEST COPY AVAILABLE

data to support one condition, i.e. increase or decrease, of the data.

Each of the five subcategories is converted to a subscore of 1 to 10 on the basis of the percentage scored. The sum of these subscores is the final score (range 0-50) on the computer simulation and was the score used in the reliability and validity studies.

Method

Subjects

The sample consisted of 75 high school students at a public school in a middle to upper middle class suburb of Atlanta, Georgia. The 38 of the subjects were male and 37 were female. All four years of the high school were represented in the sample as it was drawn from 9th grade Physical Science, Chemistry, and Chemistry II classes. The age of the sample varied from 14years 2 months to 19 years 5 months at the time of administration of the instruments.

Experimental Procedure

Validity whether predictive or concurrent is established by comparing the scores of a sample on two separate instruments. In this study the subjects were administered four instruments (1) the computer simulation,(2) the Pendulum Task,(3) the ACS-NSTA Achievement test of High School chemistry and (4) the Lawson Clasroom Test of Formal Thinking. The Pendulum task was administered and scored as described by Pauli et al. (1972). The ACS-NSTA test is a 40 question multiple-choice format test of general chemistry on the secondary level. The Lawson test is a 15 item demonstration test of formal reasoning. It involves the presentation of a task during the demonstration of actual equipment. Subjects are then presented with written questions and write their responses and reasons for making a particular choice.

The intent of the study was to compare subjects' computer simulation scores with (1) the Pendulum task, (2) the Lawson Test total score, (2) the Lawson Test Conservation subscore, (4) the Lawson Test Proportional Reasoning subscore, (5) the Lawson Test Control of Variables Subscore, (6) the Lawson Test Combinatorial Reasoning Subscore, (7) the Lawson Test Probability subscore and(8) the ACS-NSTA test. The additional question of sex bias with instruments analogous to Piagets was suggested by DeLuca (1977). Sex bias was isolated by the APA (1974)as an aspect to investigate in any instrument.

The following questions were used to formulate statistical hypotheses;

Concurrent Validity

1. Is there a relationship between the score on the computer simulation and the result of the Pendulum task?
2. Is there a relationship between the score on the computer simulation and the total score on the Lawson Classroom TEst of Formal Thinking?
3. Is there a relationship between the score on the computer simulation and any of the subscores of the Lawson test?

Predictive Validity

4. Is there a relationship between the score on the

BEST COPY AVAILABLE

computer simulation and the score on the ACS-NSTA test?

Face or Construct Validity

5. Will a panel of experts agree that the computer simulation and the associated scoring system is a theoretically consistent method by which to test for the notion of control of variables?

Test-Retest Reliability

6. Is there a relationship between the scores on the computer simulation and a readministration of the same simulation at a later time?

Statistical Analysis

Validation studies require that the scores on the instrument being investigated be compared with a criterion or an established instrument which measures the same construct. In most cases, relationships are established using correlation techniques (Amos, Brown, & Mink, 1965). The statistic used in most studies is the Pearson Product Moment correlation Coefficient. Another school of thought (Bentler, 1971; Siegel, 1956; Nagy & Griffiths, 1982) suggest that nonparametric statistic are more appropriate in comparing scores on tests of cognitive level. These authors based their assertion on the belief that any data which numerically describe cognitive level is ordinal rather interval. The nature of the data; therefore, indicates that all relationships established in this study will have to be on the basis of both types of statistics.

Once the data were collected, they were grouped for statistical analysis. To establish test-retest reliability the scores on the computer simulation for two different administration were compared with a Pearson product-moment correlation technique.

To establish concurrent validity, two criteria were used: the Lawson Classroom Test of formal Thinking and Inhelder and Piaget's Pendulum Task. The score on the computer simulation was compared with the total Lawson Test score as well as the score on conservation, control of variables, combinatorial logic, probability and proportional reasoning subscores of the Spearman Rho technique. To compare the score on the computer simulation with the score on the Pendulum Task the statistical test of choice was a one-way analysis of variance. A biserial correlation was also calculated for the relationship, since the pendulum Task results in ordinal data, while the score on the computer simulation is interval. Kubitsyn(1982)suggests that a biserial correlation is the statistic appropriate for such a relationship.

The predictive validity was established by comparing the score on the computer simulation with the score on the ACS-NSTA test if High School Chemistry, 1983 version. The scores on both instruments were compared using both a Pearson Product-moment technique and a Spearman Rho technique.

A panel of experts was consulted to establish the construct validity o the computer simulation. Six experts, who had previously published in the areas of cognitive development and instrument development were chosen. In order to exclude a gender effect, a one-way analyis of variance was used

BEST COPY AVAILABLE

exclude a gender effect, a one-way analyis of variance was used to determine the signficance of the variance of groups of male and female students on the computer simulation. The presence of a gender effect on electronic versions of the tasks was demontrated by DeLuca (1979).

Results

When the computer simulation scores were compared with the results of the Pendulum Task a significant biserial correlation was calculated (rbis=.6957;p .01). When groups of formal and concrete subjects' computer simulation scores were compared a significant F ratio (F=16.99 for 74 and 1 degrees of freedom) resulted.

The computer simulation scores were correlated with the Lawson test scores and significant correlations were found between the computer simulation scores and the Lawson total score, the control of variables subscore, and the combinatorial reasoning subscore. These correlations were all significant at the .01 level and appear in Table 1.

Insert Tables 1 & 2 Here

A one-way analysis of variance resulted in a significant F ratio for the relationship between male and female scores on the computer simulation (F=8.81;p .05). The mean of the males scores was some 4 points above that of the female counter parts.

Kubitsyn (1984) suggests that the researcher adopt a corrrelation level of r=.8 for concurrent and r=.6 for predictive validity. When one compares the parametric Pearson correlations in this study with these criteria it can be seen that the criterion of r=.8 cannot be met. The spearman correlations indicate high correlations between the computer simulation and scores on the Lawson total score, the control of variables subscore and the combinatorial reasoning subscore (rho=.9909;rho=.7918; and rho=.8845 respectively). The criteria have been met in these cases. The test-retest statistics are r=.68. Although several months had elapsed between administrations, on must be cautious in the interpretation of these statistics. The sample was a convenience sample and as such consisted of the more talented students.

There was little indication of a relationship between the computer simulation scores and the scores on the NSTA-ACS test of High School Chemistry. This may be considered to be advantageous, since it indicates that there is little influence of Chemistry knowledge on the computer simulation score.

Results indicate that the male subgroup is more likely to score higher on the computer simulation. The related question on the questionaire sent to the panel of experts indicated afffirmative responses in all cases. In addition, one panel member indicated that familiarity with physical science concepts rather than familiarity with computers may have influenced the male subjects' computer simulation scores.

The computer simulation appears to have a certain degree of construct validity. This conclusion is made on the basis of the

BEST COPY AVAILABLE

panel's comments and the statistics indicating a significant variance between concrete and formal subjects on the computer simulation score as well as on the results of the biserial correlation. Any further conclusions about the validity of the instrument must be made on the basis of the non-parametric, Spearman correlations. If one assumes cognitive level to be measured on an ordinal scale rather than on the traditional interval scale, concurrent validity has been established.

BEST COPY AVAILABLE

```
>LIST

 10  REM  HABER PROCESS PROGRAM
 20  REM  BY RICHARD A. BORST, EDS
 30  REM  GWINNETT COUNTY SCHOOLS
 40  REM  SNELLVILLE, GEORGIA 30278
 50  REM  TO BE USED IN A PIAGETIAN ANALYSIS OF STUDENT DEVELOPMENT LEVEL

100  PRINT "WHAT IS THE PRESSURE OF YOUR SYSTEM"
120  INPUT P
300 P9 =  INT ( RND (1) * 300 + 1000)
331  IF P < P9 THEN 360
340  PRINT P;"ATM IS TOO MUCH PRESSURE AN EXPLOSION MAY OCCUR"
341  PRINT : PRINT : PRINT
345  GOTO 100
360  INPUT "WHAT IS THE OPERATING TEMPERATURE";T
365 T9 =  INT ( RND (1) * 100 + 950)
370  IF T < T9 THEN 400
375  PRINT "THAT S TOO HOT - A MELTDOWN IS LIKELY"
380  GOTO 360
400 T = T + 273
410  PRINT "WANT TO USE A CATALYST (Y OR N)"
411  INPUT Z$
412  IF Z$ = "Y" THEN 430
415 E = 80000:Z$ = "NO CATALYST"
420  GOTO 550
430  PRINT "WHICH ONE IRON OR TUNGSTEN"
431  INPUT Z1$
450  IF Z1$ = "IRON" THEN E = 60000: GOTO 550
470  IF Z1$ = "TUNGSTEN" THEN E = 40000: GOTO 550
480  PRINT "PLEASE ANSWER NOW! IRON OR TUNGSTEN"
485  GOTO 430
550 K =  EXP (((-45e3 + (1.1) * P) / T) - 14)
560 B = (0.03    K) + (2 * P)
570  LET R = ((B ^ 2) - (4 * (P ^ 2))) ^ .5
580 P1 = (B - R) / 2
590 A = 100 * P1 / P
600  PRINT : PRINT : PRINT
610  PRINT "% YIELD OF AMMONIA IS "A
615  PRINT P;"    ";T;"    ";Z1$;"    "A
620  PRINT "WANT TO TRY AGAIN"
621  INPUT A$
630  IF A$ = "Y" THEN 100
700  END
```

BEST COPY AVAILABLE

Table 3

Parametric Intercorrelation Table

	Computer Simulation	Lawson Total	Lawson Conservation	Lawson Proportional Reasoning	Lawson Control of Variables	Lawson Combinatorial Reasoning	Lawson Probability
Computer Simulation	--	.4498	.1534	.1716	.5112	.34	.1187
Lawson Total		--	.6116	.5883	.8799	.7192	.5202
Lawson Conservation			--	.2207	.3469	.3711	.1078
Lawson Proportional Reasoning				--	.1841	.3065	.1538
Lawson Control of Variables					--	.6179	.3780
Lawson Combinatorial Reasoning						--	.4025
Lawson Proportional Thinking							--

Table 4

Nonparametric Correlation Table

	Computer Simulation	Lawson Total	Lawson Conservation	Lawson Proportional Reasoning	Lawson Control of Variables	Lawson Combinatorial Reasoning	Lawson Probability
Computer Simulation	--	.9009	.8253	.8495	.7918	.8845	.9806
Lawson Total		--	.1707	.1695	.4298	.4757	.7559
Lawson Conservation			--	.5035	.4009	.4856	.2699
Lawson Proportional Reasoning				--	.3401	.3716	.2575
Lawson Control of Variables					--	.6972	.2851
Lawson Combinatorial Reasoning						--	.5287
Lawson Proportional Thinking							--

BEST COPY AVAILABLE

American Psychological Association. (1974). Standards for educational and psychological tests. Washington, DC: APA Press.

Ames, J. B., Brown, F. L., & Mink, O. (1965). Statistical concepts. New York: Harper & Row.

Ary, D., Jacobs, L. C., & Razavieh, A. (1972). An introduction to research in education. Chicago: Holt, Rinehart, & Winston.

Bastien, C., & Hoyet, P. (1975). Etude genetique autonome des operations necessaires pour construire un ensemble ordonne. Cahiers de Psychologie, 18, 51-61.

Brainerd, C. J. (1978). Piaget's theory of intelligence. Englewood Cliffs, NJ: Prentice-Hall.

Bruner, J. S., Oliver, R., & Greenfield, P. (1966). Studies in cognitive growth. New York: Wiley & Sons.

Dale, L. G. (1970). The growth of systematic thinking: Replication and analysis of Piaget's first chemical experiment. Australian Journal of Psychology, 22, 277-286.

Delucia, F. P. (1979). Application and analysis of Piaget's first chemical experiment. Journal of Research in Science Teaching, 16(1), 1-11.

deVries, R. (1969). Constancy of generic identity in the years three to six. Society for Research in Child Development, 34(3, Serial No. 127).

Driver, R., & Easly, J. (19). Pupils and paradigms: A review of literature related to concept development in adolescent science students. Studies in Science Education, 5, 61-84.

Flavell, J. (1963). The developmental psychology of Jean Piaget. Princeton, NJ: D. Van Nostrand Co.

Flavell, J. (1977). Cognitive development. New York: Prentice-Hall.

Forman, G., & Sigel, I. (1979). Cognitive development: A life span view. Monterey, CA: Brooks/Cole.

Gardner, H. (1984). Assessing intelligences: A comment on testing without I.Q. tests. Phi Delta Kappan, 65, 699-700.

Gray, W. (1971). Development of a Piagetian-based written test: A criterion-referenced approach. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Guenther, W. C. (1977). Power and sample size determination when the alternative hypotheses are given in quantities. The American Statistician, 31(1), 117-118.

Huck, S. W., Cormier, W. H., & Bounds, W. G. (1974). Reading statistics and research. New York: Harper & Row.

Inhelder, B., & Piaget, J. (1958). The growth of logical thinking: From childhood to adolescence. New York: Basic Books.

Ironson, G. A., & Schwenimer, R. W. (1977). Introductory statistics for psychology. Boston: Allyn and Bacon.

Karplus, R., & Lavatelli, C. (1969). The developmental theory of Piaget: Conservation. San Francisco: Davidson Film Products.

Karplus, R. H., & Peterson, R. W. (1970). Intellectual development beyond high school. School Science and Mathematics, 70, 813-820.

Kats, R. M., & McSweeney, M. F. (1980). Errors of misclassification in neo-Piagetian research. Educational and Psychological Measurement, 40, 845-854.

BEST COPY AVAILABLE

Kubiszyn, T. (1984). Educational testing and measurement. Glenville, IL: Scott, Foresman.

Lawson, A. E. (1978). The development and validation of a test of formal reasoning. Journal of Research in Science Teaching, 15(1), 11-24.

Lawson, A. E., Nordland, F. H., & DeVito, A. (1975). A relationship of formal reasoning to achievement, aptitudes, and attitudes in preservice teachers. Journal of Research in Science Teaching, 12, 423-432.

Longeot, F. (1975). Analyse statistique de trois tests genetique collectifs. Bulletin de l'Institute National D'Etude, 20, 219-237.

Minimum, E. W. (1977). Statistical reasoning in psychology and education. New York: John Wiley & Sons.

Nagy, P., & Griffiths, A. K. (1982). Limitations of recent research relating Piaget's theory to adolescent thought. Review of Educational Research, 52, 513-556.

Nguygen-Xuan, A. (1976). Suite aux automates de seriation de Frey. Cahiers du Psychologie, 19, 101-108.

Pauli, L., Nathan, H., Droz, R., & Grize, J. B. (1977). Piagetian inventories: The experiments of Jean Piaget. Paris: Organization for Economic Co-operation and Development.

Phillips, D. G. (1982). Piaget's model of intellectual development: Deriving classroom applications. In R. E. Yager (Ed.), What research says to the science teacher. Washingon, DC: National Science Teachers Association.

Piaget, J. (1962). Play, dreams, and imitation in childhood. New York: Norton.

Piaget, J. (1964). Development and learning. In R. E. Ripple & U. N. Rockcastle (Eds.), Piaget rediscovered: A report of the Conference on Cognitive Studies and Curriculum Development. Ithaca, NY: School of Education, Cornell University.

Piaget, J. (1967). Six psychological studies. New York: Random House.

Piaget, J., Grize, J. B., Szeminska, A., & Vinh, B. (1968). Epistemologie a psychologie de la fonction (Vol. 23 of the Etudes d'epistemologie genetique series). Paris: Presses Universitaires de France.

Pulaski, M. S. (1971). Understanding Piaget: An introduction to children's cognitive development. New York: Harper & Row.

Raven, R. (1973). The development of a test of Piaget's logical operations. Science Education, 57, 377-385.

Raven, R. (1977). The development of a test of Piaget's operative comprehension. Science Education, 61, 271-278.

Shayer, M., Adey, P., & Wylam, H. (1974). Group tests of cognitive development: Ideals and a realization. Journal of Research in Science Teaching, 55, 447-458.

Siegel, S. (1956). Non-parametric statistics. New York: McGraw-Hill.

Sommerville, S. C. (1974). The pendulum problem: Patterns of performance defining developmental stages. British Journal of Educational Psychology, 44, 266-281.

Stanley, J. C. (1971). In R. L. Thorndike (Ed.), Educational measurement. Washington, DC: American Council on Education.

Staver, J. R., & Harty, H. (1979). Validation of an instrument and testing protocol for measuring the combinatorial analysis schema. Journal of Research in Science Teaching, 16, 53-60.

BEST COPY AVAILABLE